

STATISTIKA

NEPARAMETARSKI TESTOVI

Parametarski test je baziran na određenim pretpostavkama vezanim za

- parametre
- raspodelu populacije iz koje je uzorak

Validnost testova zacisi od pretpostavki

Neparametarski test ne uključuje nikakve preduslove u vezi populacije i parametara populacije iz koje je uzorak.

- Ne zahtevaju precizna "merenja" kao parametarski testovi.
- Neparametarskim se može povećati moć jednostavnim povećanjem n .

Ispituju i parametre i populaciju, razlike i slično.

Često nemamo pretpostavke o raspodeli.

Prednosti neparametarskih testova:

- 1) Iskazi o verovatnoći - tačne verovatnoće bez obzira na oblik raspodele populacije iz koje je uzorak.
 - Izvesno se pretpostavlja neprekidnost, ali to je i pretpostavka parametarskih.
- 2) Veličina uzorka mala nema alternative sem kad je raspodela populacije precizno definisana.
- 3) Nekoliko podobnih neparametarskih za uzorke iz nekoliko različitih populacija.
- 4) Pogodni za podatke koji su rangirani, kao i čije očigledne numeričke vrednosti imaju snagu ranga (ordinalna skala).
- 5) Neparametarski se mogu primeniti i na podatke u nominalnoj skali (kvalitativni, opisni podaci)
- 6) Obično mnogo lakši za učenje i primenu.

Nedostaci neparametarskih testova

- 1) Kada su sve pretpostavke parametarskog testa ispunjene u podacima neparametarski testovi često gube informacije iz podataka (parametarski su jači)
- 2) Još uvek ne postoji neparametarski metod za testiranje interakcija u modelu analize varijanse.
- 3) Zahtevaju više rada

Testiranje hipoteza na osnovu jednog uzorka - da li je uzorak magao doći iz neke poznate populacije.

- značajna razlika u lokaciji (centralna tendencija) uzorka i populacije? - m
- značajna razlika u frekvencijama? - f
- značajna razlika između posmatranih i očekivanih proporcija? - p
- da li je uzorak izvučen iz populacije određenog oblika ili forme (kao normalna, uniformna...)?

Testiranje hipoteza na osnovu dva uzorka - porede dva uzorka i ispituju da li su ta dva uzorka došla iz iste populacije.

χ^2 TEST

Test saglasnosti

Problem saglasnosti teorijskih rezultata vezanih za verovatnoću i stvarnih observacija.
Da li slučajna promenljiva ima datu raspodelu?

H_0 - obeležje X ima datu raspodelu

H_1 - obeležje X nema datu raspodelu

Postupak:

1. Skup mogućih vrednosti za X podelimo u r disjunktivnih podskupova $S_1, S_2, \dots, S_r$,
2. Označimo sa $m_k, k=1 \dots r$, broj slučajnih proenljivih iz uzorka koji upadaju u skup S_k
3. Neka je $p_k = \{X \in S_k\}$ verovatnoća da $X \in S_k$ - ove verovatnoće dobijamo iz pretpostavljene raspodele.
4. Sada koristimo statistiku (n – veličina uzorka)

$$\tau = \sum_{k=1}^r \frac{(m_k - np_k)^2}{np_k} : \chi_{r-1}^2$$

– Razlika između stvarnih i očekivanih frekfencija

5. Kritičnu vrednost određujemo iz uslova

$$P\{\tau > \chi_0^2\} = \alpha \text{ - desnostrana}$$

$$\tau \in C \Rightarrow H_0 \text{ odbacujemo}$$

$$\tau \notin C \Rightarrow H_0 \text{ prihvatamo}$$

Posmatra se razlika između očekivanih frekvencija iz raspodele i realizovanih u uzorku, pa se na osnovu njih formira χ^2 statistika

Test podobnosti modela

Potpuno poznavanje raspodele retko u praksi. Češći izvestan broj nepoznatih parametara.

Sve isto samo što $p_k = \{X \in S_k\}$ verovatnoća da $X \in S_k$ zavisi od nepoznatih parametara.

$$\tau = \sum_{k=1}^r \frac{(m_k - np_k)^2}{np_k} : \chi_{r-l-1}^2$$

Korigovani stepeni slobode – umanjani za broj nepoznatih parametara
l - broj nepoznatih parametara koje treba oceniti na osnovu uzorka

Tabela kontingencije – hipoteza o nezavisnosti obeležja X i Y

$$\tau = n \sum_{i=1}^r \sum_{j=1}^s \frac{\left(n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n} \right)^2}{n_{i.} \cdot n_{.j}} : \chi^2_{(r-1)(s-1)}$$

Da li se empirijski dobijene frekvencije razlikuju od očekivanih

ZADACI:

1. Na osnovu 120 merenja vrednosti slučajne promenljive X dobijeni su sledeći podaci:

X	0	1	2	3	4
f	8	26	50	27	9

Sa pragom značajnosti 0.01, proveriti pretpostavku da X ima Binomnu raspodelu sa varijansom jednakom 1.

Rešenje:

$$X : B(n, p) ? \quad X : \{0, 1, 2, \dots, n\}$$

H_0 : „X ima $B(n; p)$ raspodelu“

H_1 : „X nema $B(n; p)$ raspodelu“

$$n=4 \quad n=120 \text{ (obim uzorka)} \quad r=5 \quad p=?$$

$$\sigma^2 = npq = 1$$

$$q = 1 - p$$

$$1 = 4p(1 - p)$$

$$4p^2 - 4p + 1 = 0$$

$$p_{1/2} = \frac{4 \pm \sqrt{16 - 16}}{8} \Rightarrow p = 0.5$$

X	0	1	2	3	4
m_k	8	26	50	27	9
p_k	0.0625	0.25	0.375	0.25	0.0625

$$\tau = \sum_{k=1}^r \frac{(m_k - np_k)^2}{n \cdot p_k} : \chi^2_{r-1}$$

- U slučaju potpune podudarnosti očekivane vrednosti sa empirijskom (dobijenom) vrednost statistike biće 0; postoje slučajne fluktuacije u uzorku
- Pitanje koje se postavlja i koje se proverava je da li razlika u empirijskim i očekivanim frekvencijama značajno odstupa od 0; da li postoji značajna razlika između empirijskog i očekivanog

$$p_k = P(X = k) = \binom{n}{k} p^k q^{n-k}, k = 0, 1, 2, 3, 4. \quad p_2 = \binom{4}{2} 0.5^2 \cdot 0.5^{4-2} = 0.375$$

$$p_0 = \binom{4}{0} 0.5^0 \cdot 0.5^{4-0} = 0.0625 \quad p_3 = \binom{4}{3} 0.5^3 \cdot 0.5^{4-3} = 0.25$$

$$p_1 = \binom{4}{1} 0.5^1 \cdot 0.5^{4-1} = 0.25 \quad p_4 = \binom{4}{4} 0.5^4 \cdot 0.5^{4-4} = 0.0625$$

$$\tau = \sum_{k=0}^4 \frac{(m_k - np_k)^2}{n \cdot p_k} = \frac{(8 - 120 \cdot 0.0625)^2}{120 \cdot 0.0625} + \frac{(26 - 120 \cdot 0.25)^2}{120 \cdot 0.25} + \frac{(50 - 120 \cdot 0.375)^2}{120 \cdot 0.375} + \frac{(27 - 120 \cdot 0.25)^2}{120 \cdot 0.25} + \frac{(9 - 120 \cdot 0.0625)^2}{120 \cdot 0.0625} = 1.722$$

mao

$$P\{\tau > \chi_0^2\} = 0.01 \quad P\{\tau < \chi_0^2\} = 0.99 \Rightarrow \chi_0^2 = 13.272$$

χ_4^2

$$c = (13.272; +\infty) \quad \tau \notin c \Rightarrow H_0 \text{ prihvataemo, tj. zaključujemo da } X \text{ ima } B(4; 0.5).$$

2. Ispitati saglasnost podataka datih uzorkom sa Normalnom raspodelom, sa nivoom značajnosti od 5%

X	(1,3)	(3,5)	(5,7)	(7,9)	(9,11)	(11,13)
m _k	5	16	30	29	13	7

Rešenje:

H_0 : "X ima $N(m, \sigma^2)$ raspodelu"

H_1 : "X nema $N(m, \sigma^2)$ raspodelu"

$$\tau = \sum_{k=1}^r \frac{(m_k - np_k)^2}{n \cdot p_k} : \chi_{r-l-1}^2$$

$$n = \sum_k f_k = 100 \quad r = 6 \quad l = 2 \quad \text{jer ocenjujemo } m \text{ i } \sigma^2$$

$$\hat{m} = \bar{x} = \frac{1}{n} \sum_{k=1}^r x_k \cdot m_k = \frac{1}{100} (2 \cdot 5 + 4 \cdot 16 + 6 \cdot 30 + 8 \cdot 29 + 10 \cdot 13 + 12 \cdot 7) = 7$$

$$\hat{\sigma}^2 = S^2 = 6.2 \quad \hat{\sigma} = 2.5$$

Određujemo nepoznate verovatnoće p_k standardizacijom slučajne promenljive X , tj. preko:

$$X^* = \frac{X - \hat{m}}{\hat{\sigma}} : N(0,1)$$

$$p_1 = P\{1 < X < 3\} = P\left\{\frac{1-7}{2.5} < X^* < \frac{3-7}{2.5}\right\} = P\{-2.4 < X^* < -1.6\} = \Phi(-1.6) - \Phi(-2.4) = 1 - \Phi(1.6) - (1 - \Phi(2.4)) = 1 - 0.9452 - (1 - 0.9918) = 0.05$$

$$p_2 = P\{3 < X < 5\} = P\left\{\frac{3-7}{2.5} < X^* < \frac{5-7}{2.5}\right\} = 0.16$$

$$p_3 = P\{5 < X < 7\} = P\left\{\frac{5-7}{2.5} < X^* < \frac{7-7}{2.5}\right\} = 0.29$$

$$p_4 = P\{7 < X < 9\} = P\left\{\frac{7-7}{2.5} < X^* < \frac{9-7}{2.5}\right\} = 0.29$$

$$p_5 = P\{9 < X < 11\} = P\left\{\frac{9-7}{2.5} < X^* < \frac{11-7}{2.5}\right\} = 0.16$$

$$p_6 = P\{11 < X < 13\} = P\left\{\frac{11-7}{2.5} < X^* < \frac{13-7}{2.5}\right\} = 0.05$$

$$\tau = \frac{(5-100 \cdot 0.05)^2}{100 \cdot 0.05} + \frac{(16-100 \cdot 0.16)^2}{100 \cdot 0.16} + \dots + \frac{(7-100 \cdot 0.05)^2}{100 \cdot 0.05} = 1.4$$

$$P(\tau > \chi_0^2) = 0.05 \quad P(\tau < \chi_0^2) = 0.95 \quad \Rightarrow \quad \chi_0^2 = 7.815$$

χ_{6-2-1}^2

$$\tau \notin c = (7.815; +\infty) \Rightarrow H_0 \text{ prihvatamo}$$

3. Od 150 bacanja kockice pali su brojevi:

1	2	3	4	5	6
24	20	20	23	33	30

Testirati hipotezu o homogeneosti kockice (Sve strane iste verovatnoće).

Rešenje:

X : strana na kockici

H_0 : "X ima ravnomernu raspodelu - svaka strana kocke je jednako verovatna"

H_1 : "X nema ravnomernu raspodelu"

$$r = 6 \quad n = 150 \quad l = 0 \quad \alpha = 0,05$$

$$\tau = \sum_{k=1}^r \frac{(m_k - np_k)^2}{n \cdot p_k} : \chi_{r-1}^2$$

$$\tau = \frac{\left(24 - 150 \cdot \frac{1}{6}\right)^2}{150 \cdot \frac{1}{6}} + \dots + \frac{\left(30 - 150 \cdot \frac{1}{6}\right)^2}{150 \cdot \frac{1}{6}} = 5.76$$

$$P\{\tau > \chi_0^2\} = 0.05 \quad P\{\tau < \chi_0^2\} = 0.95 \quad \Rightarrow \quad \chi_0^2 = 11.070$$

χ_5^2

$$c = (11.070; +\infty) \quad \tau \notin c \Rightarrow H_0 \text{ prihvatamo}$$

4. Testira se zavisnost odgovora 200 ispitanih kandidata na pitanje: X : "Da li pušite?", Y: "Da li je pušenje štetno?". Sa rizikom od 1% proveriti da li postoji zavisnost između odgovora, ako su dobijeni sledeći rezultati:

	Y	"da"	"ne"
X			
"da"		80	80
"ne"		25	15

Rešenje:

H_0 { "X i Y su nezavisni odgovori" }

H_1 { "X i Y nisu nezavisni odgovori" }

Koristimo statistiku:

$$\tau = n \sum_{i=1}^r \sum_{j=1}^s \frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{n} \right)^2}{n_{i.} n_{.j}} : \chi_{(r-1)(s-1)}^2$$

	Y	"da"	"ne"	$n_{i.}$
X				
"da"		80	80	160
"ne"		25	15	40
$n_{.j}$		105	95	200

- Kada su zajedničke frekvencije jednake proizvodu marginalnih, obeležja su nezavisna

$$\tau = 200 \left[\frac{\left(80 - \frac{160 \cdot 105}{200} \right)^2}{160 \cdot 105} + \frac{\left(80 - \frac{160 \cdot 95}{200} \right)^2}{160 \cdot 95} + \frac{\left(25 - \frac{105 \cdot 40}{200} \right)^2}{105 \cdot 40} + \frac{\left(15 - \frac{95 \cdot 40}{200} \right)^2}{95 \cdot 40} \right] = 2$$

$$P\{\tau > \chi_0^2\} = 0.01 \quad P\{\tau < \chi_0^2\} = 0.99 \quad \Rightarrow \quad \chi_0^2 = 6.63$$

$\chi_{(2-1)(2-1)}^2$

$$c = (6.63; +\infty) \quad \tau \notin c \quad H_0 \text{ prihvatamo.}$$

Test koraka

Testiranje slučajnosti uzorka

H_0 – “Uzorak je slučajan”

Izbor kritične oblasti zavisi od alternativne hipoteze

- a. H_1 “slučajnost uzorka je narušena zbog monotonosti očekivane vrednosti u populaciji”
Levostrana kritična oblast $P\{\tau < z_0\} = \alpha$
- b. H_1 “slučajnost je narušena zbog periodičnih promena očekivane vrednosti u populaciji”
Desnostrana kritična oblast $P\{\tau > z_0\} = \alpha$
- c. H_1 “uzorak nije slučajan”
Obostrana kritična oblast $P\{|\tau| > z_0\} = \alpha$

$$\tau = \frac{K - E(K)}{\sqrt{Var(K)}} : N(0,1)$$

$$E(K) = \frac{n+2}{2} \quad Var(K) = \frac{n(n-2)}{4(n-1)}$$

K - broj koraka

Testiranje jednakosti raspodela

H_0 “oba uzorka su iz populacije sa istom raspodelom” ili
 H_0 “X i Y imaju istu raspodelu”

H_1 “raspodele nisu iste” ili
 H_1 “X i Y nemaju istu raspodelu”

$$\tau = \frac{K - E(K)}{\sqrt{Var(K)}} : N(0,1)$$

$$E(K) = \frac{2n_1n_2}{2} + 1 = \mu \quad Var(K) = \frac{(\mu-1)(\mu-2)}{n-1}$$

$$n = n_1 + n_2$$

Kritična oblast je levostrana $P\{\tau < z_0\} = \alpha$

5. Izmerena su odstupanja u težini jednog proizvoda u gramima.

12, 11, 7, 10, 6, 4, 18, 11, 8, 9

Da li je posmatrani uzorak slučajan?

Rešenje:

H_0 : "Uzorak je slučajan"

H_1 : "Uzorak nije slučajan"

$$\tau = \frac{K - \frac{n+2}{2}}{\sqrt{\frac{n(n-2)}{4(n-1)}}} : N(0,1)$$

$n = 10$

Postupak za određivanje broja koraka K:

1^o Elemente uzorka poredamo u neopadajući niz X'

4, 6, 7, 8, 9, 10, 11, 11, 12, 18

2^o Odredimo medijanu po formuli:

$$Me = \begin{cases} X'_{\frac{n+1}{2}} & : n - \text{нечетно} \\ \frac{1}{2} \left(X'_{\frac{n}{2}} + X'_{\frac{n}{2}+1} \right) & : n - \text{четно} \end{cases}$$

$$M_e = \frac{1}{2} (x'_5 + x'_6) = \frac{1}{2} (9 + 10) = 9.5$$

3^o Označimo sa $Y = \begin{cases} - & : \text{ако је } X < M_e \\ + & : \text{ако је } X \geq M_e \end{cases}$

Vrednosti u uzorku zamenimo ovako definisanim vrednostima Y. U slučaju podatka iz ovog zadatka, odgovarajući niz Y je:

+ + - + - + + - -

4^o Korak je broj nizova uzastopnih istih znakova. U našem primeru $K = 6$

$$\tau = \frac{K - \frac{n+2}{2}}{\sqrt{\frac{n(n-2)}{4(n-1)}}} = \frac{6 - \frac{10+2}{2}}{\sqrt{\frac{10(10-2)}{4(10-1)}}} = 0$$

$$P\{|\tau| > z_0\} = \alpha = 0.05 \quad P\{|\tau| < z_0\} = 0.95$$

tab

$$2\Phi(z_0) - 1 = 0.95 \Rightarrow z_0 = 1.96$$

$$C = (-\infty; -1.96) \cup (1.96; +\infty)$$

N

$\tau \notin c \Rightarrow H_0$ prihvataмо i zaključujemo da je uzorak slučajan.

6. Ispituje se uzorak iz dve populacije A i B. Iz populacije A je izvučen uzorak X:
4, 2, 1, 12, 2, 3, 9, 3, 4, 3, 5, 12
Iz populacije B je izvučen uzorak Y:
12, 1, 2, 8, 4, 7, 13, 2, 15, 4, 11, 2, 17, 6, 20
Sa pragom značajnosti 0.05 testirati hipotezu da populacije A i B imaju istu raspodelu.

Rešenje:

H_0 : "A i B imaju istu raspodelu";
 H_1 : "A i B nemaju istu raspodelu"

$$\tau = \frac{K - E(K)}{\sqrt{\text{var}(K)}} : N(0,1)$$

$$E(K) = \mu = \frac{2n_1n_2}{n} + 1 = \frac{2 \cdot 12 \cdot 15}{12 + 15} + 1 = 14.33$$

$$\text{Var}(K) = \frac{(\mu - 1)(\mu - 2)}{n - 1} = \frac{(14.33 - 1)(14.33 - 2)}{(12 + 15) - 1} = 6.32$$

Određujemo broj koraka K:

Poređamo vrednosti uzorka X i Y po veličini

x3 y2 x2 x5 y12 y3 y8 x10 x6 x8 x9 x1 y5 y10 x11 y14 y6 y4 x7 y11x4 x12 y9 y7 y15 y1 y13

Broj koraka K je broj nizova uzastopnih x - ova ili y - ona, tj. $K = 12$

$$\tau = \frac{12 - 14.33}{\sqrt{6.32}} = -0.92$$

$$P\{\tau < z_0\} = 0.05 \quad \Phi(+z_0) = 0.05$$

tab

$$\Phi(-z_0) = 0.95 \Rightarrow -z_0 = 1.64 \Rightarrow z_0 = -1.64$$

N

$$C = (-\infty; -1.64)$$

$\tau \notin C \Rightarrow H_0$ prihvatamo

W test – Mann-witney

Spada u klasu rang testova

H_0 : “Oba uzorka su iz iste populacije” ili
 H_0 : “Obe populacije imaju istu raspodelu”

H_1 : “Uzorci nisu iz iste populacije”

$$\tau = \frac{W - \frac{n_1 n_2}{2}}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)}} \sqrt{12} : N(0;1)$$

W – porede se vrednosti I i II uzorka. Za svaki element I uzorka određuje se broj “inverzija” – broj elemenata II uzorka koji su manji od te vrednosti. Najbolje je napraviti rang listu zajedničkih elemenata.

$n_1 > 4$
 $n_2 > 4$
 $n_1 + n_2 > 20$

Kritična oblast je levostrana $P\{\tau < w_0\} = \alpha$

7. Radi upoređivanja visine košarkaša u dve lige, izabrano je, na slučajan način, 11 košarkaša iz prve i 12 košarkaša iz druge lige i dobijeni su podaci
Prva liga: 1.98 2.31 2.25 2.07 1.89 2.13 2.22 2.01 1.86 1.95 1.84
Druga liga: 2.19 2.26 2.28 1.90 2.03 2.08 2.00 2.04 2.32 2.02 2.24 2.35
W-testom ispitati da li postoji razlika između visina košarkaša u ove dve lige, sa pragom značajnosti 0,01.

Rešenje:

H_0 : “Oba uzorka su iz iste populacije”
 H_1 : “Uzorci nisu iz iste populacije”

$n_1 = 11 > 4$
 $n_2 = 12 > 4$
 $n_1 + n_2 = 23 > 20$

$$\tau = \frac{W - \frac{n_1 n_2}{2}}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)}} \sqrt{12} : N(0;1)$$

Poređaćemo po veličini sve vrednosti u oba uzorka i rangirati ih.

Rang	Vrednost	Uzorak
1	1.84	1
2	1.86	1
3	1.89	1
4	1.90	2
5	1.95	1
6	1.98	1
7	2.00	2
8	2.01	1
9	2.02	2
10	2.03	2
11	2.04	2
12	2.07	1
13	2.08	2
14	2.13	1
15	2.19	2
16	2.22	1
17	2.24	2
18	2.25	1
19	2.26	2
20	2.28	2
21	2.31	1
22	2.32	2
23	2.35	2

$$W = 0 + 0 + 0 + 1 + 1 + 2 + 5 + 6 + 7 + 8 + 10 = 40$$

$$\tau = \frac{40 - \frac{11 \cdot 12}{2}}{\sqrt{11 \cdot 12 (11 + 12 + 1)}} \sqrt{12} = \frac{40 - 66}{\sqrt{264}} = -1.6$$

$$P\{\tau < w_0\} = \alpha = 0.01 \quad \overset{tab}{\Phi(-w_0) = 0.99} \Rightarrow -w_0 = 2.35 \Rightarrow w_0 = -2.35$$

N

$$C = (-\infty; -2.35)$$

$\tau \notin C$, prihvatamo hipotezu H_0 i zaključujemo da uzorci potiču iz iste populacije.

Nerešenim opservacijama sa istom vrednošću dodeljuje se prosek nerešenih vezanih opservacija

Domaći:

8. Prosečan broj grešaka u radu mašine bio je 8. Izvršena su merenja i dobijeni su rezultati:

Broj grešaka	Broj merenja
0-5	60
6-10	20
11-15	10
16-20	5
21-24	5

Da li se može smatrati da broj grešaka ima Binomnu raspodelu?

Rešenje:

$$H_0: "X \text{ ima } B(n; p)$$

$$H_1: "X \text{ nema } B(n; p)$$

$$n = 100 - \text{obim uzorka} \quad n = 24 \quad E(X) = n \cdot p \quad 24 \cdot p = 8 \quad p = \frac{1}{3} \quad B\left(24; \frac{1}{3}\right)$$

$$\tau = \sum_{k=1}^r \frac{(m_k - n \cdot p_k)^2}{n \cdot p_k} : \chi_{r-1}^2$$

Nemamo nepoznatih parametara (p nije nepoznato dobili smo ga na populaciji jer je vrednost 8 iz populacije) pa je $l=0$

Moguće vrednosti za X su: $X = \{0, 1, 2, \dots, 24\}$

U podacima je ovaj skup podeljen na podskupove:

$$S_1 = \{0, 1, 2, 3, 4, 5\}$$

$$S_2 = \{6, 7, 8, 9, 10\}$$

$$S_3 = \{11, 12, 13, 14, 15\}$$

$$S_4 = \{16, 17, 18, 19, 20\}$$

$$S_5 = \{21, 22, 23, 24\}$$

Za ove skupove odredićemo odgovarajuće verovatnoće, na osnovu pretpostavke da H ima binomnu raspodelu sa navedenim perametrima.

$$p_1 = P\{X \in S_1\} = \sum_{i=0}^5 \binom{24}{i} \cdot \left(\frac{1}{3}\right)^i \cdot \left(\frac{2}{3}\right)^{24-i}$$

$$p_2 = P\{X \in S_2\} = \sum_{i=6}^{10} \binom{24}{i} \cdot \left(\frac{1}{3}\right)^i \cdot \left(\frac{2}{3}\right)^{24-i}$$

$$p_3 = P\{X \in S_3\} = \sum_{i=11}^{15} \binom{24}{i} \cdot \left(\frac{1}{3}\right)^i \cdot \left(\frac{2}{3}\right)^{24-i}$$

$$p_4 = P\{X \in S_4\} = \sum_{i=16}^{20} \binom{24}{i} \cdot \left(\frac{1}{3}\right)^i \cdot \left(\frac{2}{3}\right)^{24-i}$$

$$p_5 = P\{X \in S_5\} = \sum_{i=21}^{24} \binom{24}{i} \cdot \left(\frac{1}{3}\right)^i \cdot \left(\frac{2}{3}\right)^{24-i}$$

Ove verovatnoće ćemo tražiti aproksimacijom. Pošto je $np = 8 < 10$, aproksimaciju ćemo izvršiti preko Puasonove raspodele sa parametrom $\lambda = 8$

Iz tablice za zbrove Puasonovih verovatnoća imamo

$$p_1 = P\{X \in S_1\} \approx \sum_{i=0}^5 e^{-8} \cdot \frac{8^i}{i!} = 0.191$$

$$p_2 = P\{X \in S_2\} \approx \sum_{i=0}^{10} e^{-8} \cdot \frac{8^i}{i!} - \sum_{i=0}^5 e^{-8} \cdot \frac{8^i}{i!} = 0.625$$

$$p_3 = P\{X \in S_3\} \approx \sum_{i=0}^{15} e^{-8} \cdot \frac{8^i}{i!} - \sum_{i=0}^{10} e^{-8} \cdot \frac{8^i}{i!} = 0.1758$$

$$p_4 = P\{X \in S_4\} \approx \sum_{i=0}^{20} e^{-8} \cdot \frac{8^i}{i!} - \sum_{i=0}^{15} e^{-8} \cdot \frac{8^i}{i!} = 0.0081$$

$$p_5 = P\{X \in S_5\} \approx \sum_{i=0}^{24} e^{-8} \cdot \frac{8^i}{i!} - \sum_{i=0}^{20} e^{-8} \cdot \frac{8^i}{i!} = 0.0001$$

$$\tau = \frac{(60 - 100 \cdot 0.191)^2}{100 \cdot 0.191} + \frac{(20 - 100 \cdot 0.625)^2}{100 \cdot 0.625} + \frac{(10 - 100 \cdot 0.175)^2}{100 \cdot 0.175} + \frac{(5 - 100 \cdot 0.0081)^2}{100 \cdot 0.0081} + \frac{(5 - 100 \cdot 0.0001)^2}{100 \cdot 0.0001} = 2631.38$$

$$P\{\tau > \chi_0^2\} = 0.05 \quad P\{\tau < \chi_0^2\} = 0.95 \quad \begin{matrix} tab \\ \Rightarrow \end{matrix} \quad \chi_0^2 = 9.488 \quad c = (9.488; +\infty)$$

$\tau \in C$ pa sa pragom značajnosti 0.01 odbacujemo pretpostavku H_0 , tj. zaključujemo da X nema Binomnu raspodelu sa parametrima $n = 24$ i $p = \frac{1}{3}$

9. Tri grupe studenata koji su u tri grupe za slušanje pohađali vežbe iz matematike pokazali su sledeći uspeh na ispitu

grupa ocene	A	B	C	uk
10	0	2	0	2
9	2	1	3	6
8	11	7	10	28
7	8	15	0	23
6	24	25	26	75
5	15	25	26	55
uk	60	75	65	200

Rešenje:

$H_0 \{X \text{ i } Y \text{ su nezavisne}\}$

$H_1 \{X \text{ i } Y \text{ nisu nezavisne}\}$

$$\tau = n \sum_{i=1}^r \sum_{j=1}^s \frac{\left(n_{ij} - \frac{n_{i.} \cdot n_{.j}}{n} \right)^2}{n_{i.} \cdot n_{.j}} : \chi_{(r-1)(s-1)}^2$$

$$\tau = 200 \left[\frac{\left(0 - \frac{60 \cdot 2}{200}\right)^2}{60 \cdot 2} + \frac{\left(2 - \frac{75 \cdot 2}{200}\right)^2}{75 \cdot 2} + \dots + \frac{\left(26 - \frac{65 \cdot 66}{200}\right)^2}{65 \cdot 66} \right] = 21,76$$

$$P\{\tau > \chi_0^2\} = 0.05 \quad P\{\tau < \chi_0^2\} = 0.95 \quad \Rightarrow \quad \chi_0^2 = 18.307$$

$$\chi_{(3-1)(6-1)}^2$$

$$c = (18.307; +\infty) \quad \tau \in c \quad \Rightarrow \quad H_0 \text{ odbacujemo}$$

10. Izabrana je grupa od 20 studenata i zabeležen je njihov broj osvojenih poena iz statistike

61 94 42 23 100 35 59 46 32 69 45 91 98 33 80 27 60 74 55 17

Sa pragom značajnosti 0.05 testirati hipotezu da je uzorak slučajan protiv alternativne da postoji određena periodičnost u pojavljivanju broja poena.

Rešenje:

17 23 27 32 33 35 42 45 46 55 59 60 61 69 74 80 91 94 98 100

$$M_e = \frac{1}{2}(x'_{10} + x'_{11}) = \frac{1}{2}(55 + 59) = 57$$

- - + + - - + + - - + + - - + + K=14

$$\tau = \frac{K - \frac{n+2}{2}}{\sqrt{\frac{n(n-2)}{4(n-1)}}} = \frac{14 - \frac{20+2}{2}}{\sqrt{\frac{20(20-2)}{4(20-1)}}} = 1.37$$

tab

$$P\{\tau > z_0\} = 0.05 \quad \Phi(z_0) = 95 \Rightarrow z_0 = 1.64 \quad C = (1.64; +\infty)$$

$$N$$

$\tau \notin c \Rightarrow H_0$ prihvatamo i zaključujemo da je uzorak slučajan.